Online Supplement: Measuring Exercise Capacity and Physical Function in Older Mice

Table of Contents

eMethods:

Physical Functional Tests: Precise methodology for the following tests has been previously published (1-5). For the determination of CFAB we used: *Treadmill (endurance capacity); Grip Test (forelimb strength); Inverted Cling (four limb strength/endurance); Rotarod (balance, coordination, gait speed, power generation); and Voluntary Wheel Running (volitional exercise capacity and activity rate.* A brief description of each procedure follows:

Rotarod (overall motor function): A Panlab LE8205 rotarod measured overall motor function (balance, coordination, stamina, power). Mice were acclimated over 2 practice sessions (1 session/day with 3 trials of varied protocols/session), followed by the testing day (3 trials, accelerating 4 to 40 rpm over 5 minutes), the outcome measure was the best time out of three trials (latency to fall).

Grip Test (forelimb strength): Using a Bioseb GT3 model grip strength tester (trapeze-style grip), we report the best of 5 trials. For one trial, the mouse was held by the tail and placed, gently, so that its forepaws grabbed the bar. Then, the mouse was smoothly pulled backed until it released the bar and the grip in Newtons (N) was recorded. There was no acclimation period, since they can learn quickly to just let go, once they realize they are not in danger if the test is repeated too often.

Treadmill (endurance / volitional fatigue): We used a PanLab LE8710 treadmill to measure volitional endurance capacity. Mice acclimated to the device with two training sessions (day 1: walking for 2 minutes maximum at 6 cm/s, day two: 1 trial walking for 2 minutes maximum at 6 cm/s, and trial 2 is the same as the test day protocol but with a maximum of 2 minutes). Mice learn quite readily to avoid the shock grid during the acclimation period. On day three (test day) the treadmill begins at 6 cm/s and accelerates 1cm/sec/20 seconds until the shock grid is touched three times.

Inverted Cling (overall strength / endurance): Mice were suspended on an inverted grid set over a padded floor. Latency to fall is measured in seconds, and the best of two trials is reported. This test relies on the innate fear that mice have of falling. Mice were not acclimated to this test as it is best done naively with limited repeats, as our experience has shown that mice can learn to just let go, since there is not a consequence (they only fall a short distance). However, we do repeat the test immediately if a mouse falls in less than 10 seconds to make sure it is not a fluke.

Voluntary Wheel Running (activity and volitional exercise rate): Mice were singly housed for one week in a cage containing a running wheel with a computerized magnetic revolution counter (Columbus Instruments). The number of revolutions of the wheel were counted and converted into the outcome measure: km/day.

OTHER TESTS NOT PART OF **CFAB** TO MEASURE MUSCLE PERFORMANCE

in vitro and *in vivo* Contractile Physiology: To determine muscle physiological performance, a subset of mice were tested for dorsiflexion (mainly TA, or tibialis anterior) torque output (using *in vivo* contractile physiology), as well as the force/specific force (*in vitro* contractile physiology) of the EDL (extensor digitorum longus) and soleus (SOL) muscles of the mice. Methodology for these tests has been previously published (2-4). We also report the physiological cross-sectional area (derived as previously established) (2). These measurements **were not i**ncluded in the CFAB score. See below for further details about the methodology:

To measure muscle contraction, we used an Aurora Scientific 1300a whole mouse muscle physiology suite (model 6650LR force transducer, dual-mode lever system, hi-power biphase stimulator, signal interface, 809c in situ mouse apparatus; and software: Dynamic Muscle Control version 5.500 and Dynamic Muscle Analysis version). A brief description of each procedure follows:

in vivo Contractile Physiology (dorsiflexor torque): The mouse was anesthetized with isoflurane, set on the heated (35°C) platform, the leg was mounted into a clamp, needle electrodes were set, proper current flow was determined by eliciting a twitch (pulse duration 200 microseconds) with the stimulus increased incrementally until the maximum twitch force was recorded. This minimum current was then maintained throughout a force-frequency curve to find the maximum torque output We report maximum torque/gram body mass as the outcome measure. This is a repeatable, non-invasive, non-terminal procedure that is performed on the intact living mouse which then awakens and has no consequences as a result of the procedure. Further details can be found in previous work (4).

ex vivo / in vitro Contractile Physiology (SOL and EDL maximum isometric force): SOL and EDL muscles were cautiously removed and then perfused (95% O_2/5% CO_2) in 25°C Krebs-Ringer buffer. Suture line (#4 gauge) was tied at the myotendinous junctions, with the origin side then attached to a force transducer and the insertion (distal)

side to a static clamp. The muscles were suspended between two platinum electrodes and electric pulses stimulated the muscle to determine optimal sarcomere length (L_0), peak twitch force (P_t), and peak isometric force (P_0) as previously detailed. This is a terminal invasive procedure. Further details can be found in previous work (2,3).

Statistics:

Test	Age Group	skew	Skew SE	Skew Normal.	kurtosis	kurtosis SE	kurtosis Normal.	Shapiro-Wilkes (sig.)
Rotarod, sec.	6-month	0.54	0.46	1.19	0.07	0.89	0.08	0.379
	24-month	0.88	0.46	1.89	2.29	0.90	**2.54**	0.085
	28+-month	0.01	0.48	0.02	-0.14	0.94	-0.14	1.000
Treadmill, sec.	6-month	0.50	0.46	1.11	0.66	0.89	0.74	0.325
	24-month	0.66	0.46	1.42	0.89	0.90	0.98	0.471
	28+-month	0.31	0.48	0.64	-1.37	0.94	-1.47	**0.018**
Grip Test, N	6-month	0.35	0.46	0.77	0.36	0.90	0.40	0.397
	24-month	-0.78	0.46	-1.69	0.30	0.90	0.33	0.087
	28+-month	-0.09	0.48	-0.19	-0.93	0.94	-0.99	0.135
Inverted Cling, sec.	6-month	1.79	0.46	**3.91**	3.86	0.89	**4.36**	**0.001**
	24-month	1.07	0.46	**2.31**	1.63	0.90	1.81	0.054
	28+-month	1.98	0.48	**4.12**	4.21	0.94	**4.50**	**0.001**
Voluntary wheel Running, km/day	6-month	0.74	0.46	1.61	0.62	0.90	0.69	0.143
	24-month	0.68	0.46	1.46	0.21	0.90	0.23	0.372
	28+-month	0.68	0.48	1.41	0.98	0.94	1.05	**0.013**
CFAB original	6-month	0.37	0.46	0.80	-0.72	0.89	-0.81	0.513
	24-month	0.00	0.46	-0.01	-0.95	0.90	-1.05	0.640
	28+-month	0.19	0.48	0.39	-0.30	0.94	-0.33	0.938
Inverted Cling, logsec.	6-month	0.39	0.43	0.89	-0.51	0.85	-0.61	0.584
	24-month	-0.56	0.46	-1.21	-0.16	0.90	-0.18	0.344
	28+-month	0.19	0.46	0.41	-1.11	0.90	-1.23	0.213
CFAB w/logcling	6-month	0.32	0.46	0.70	-0.72	0.89	-0.82	0.466
	24-month	-0.08	0.46	-0.17	-1.05	0.90	-1.17	0.396
	28+-month	0.06	0.46	0.12	-1.02	0.90	-1.13	0.501

Table e1 Normality Statistics for CFAB Determinants: Abbreviations: sec=seconds, N=Newtons, km=kilometer, logsec=log10 transformation of Inverted Cling seconds, CFAB=Comprehensive Functional Assessment Battery, normal=normalized, sig=significance (P<0.05). Notes: **Bold** typeface = potential violations of normality assumptions

eResults:

CFAB:

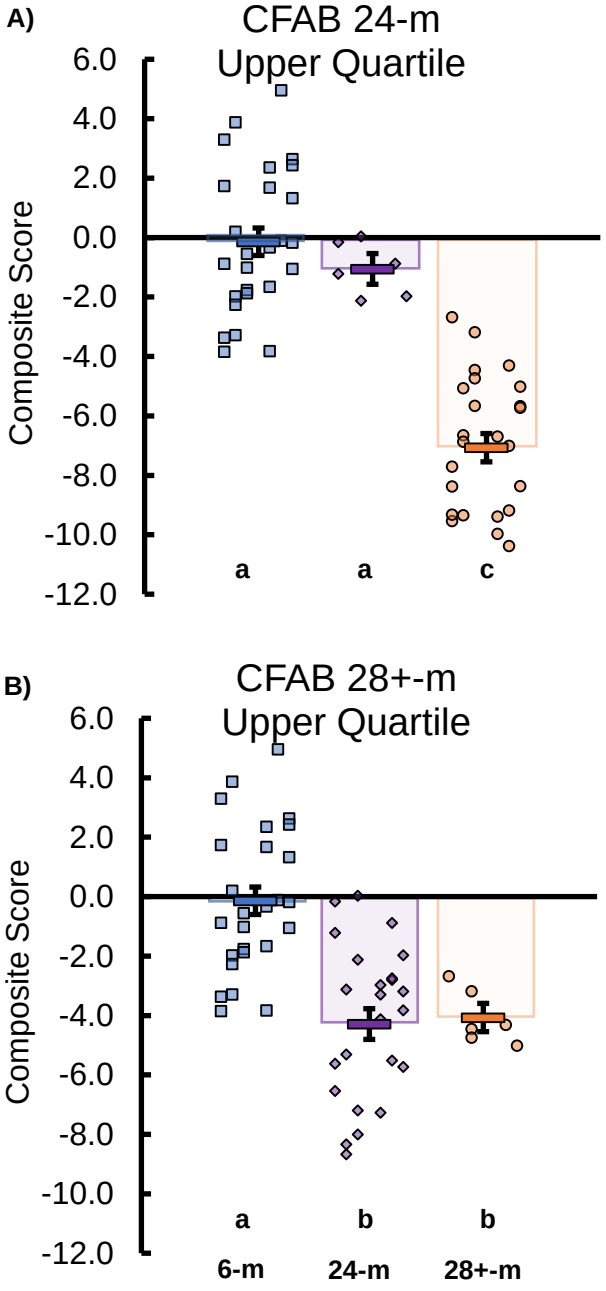

Figure e1 CFAB Upper Quartile Comparisons A) CFAB 24-month and 28+-month Upper Quartiles. 24-month mice in the upper quartile (top 25%) were not significantly different in physical function than the mean of the 6-month mice. The upper quartile of 28+-month mice performed significantly worse than both the mean 6-month mice and the upper quartile of 24-month mice ($p<0.001$). **B) CFAB 28+-month Upper Quartile.** The 28+-month mice in the upper quartile maintained their performance relative to the mean 24-month mice ($p>0.10$). **KEY:** Different letters indicate significant difference ($p<0.05$) between groups (from ANOVA and using LSD posthoc test). Each symbol [squares for 6-months, diamonds for 24-month, and circles for 28+-month old mice] indicates the CFAB of an individual mouse. The rectangles with error symbols (±SEM) in each grouping of ages indicates the mean value for that group.

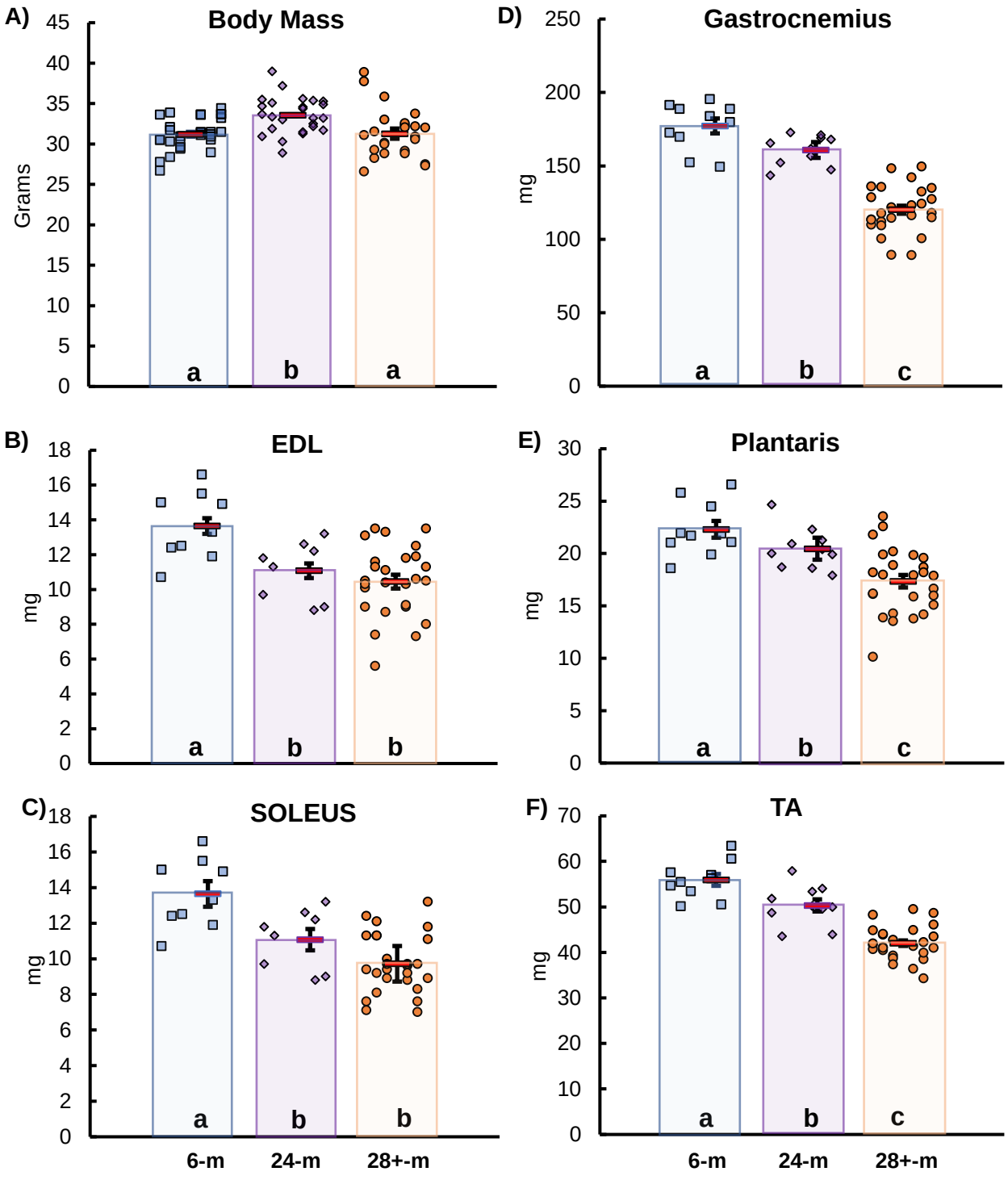

Figure e2 Body and Muscle Mass. A) Body Mass. B) Extensor Digitorum Longus (EDL) C) Soleus (SOL) D) Gastrocnemius (GAS) E) Plantaris F) Tibialis Anterior (TA) KEY: Different letters indicate significant difference ($p<0.05$) between groups (from ANCOVA adjusted for body mass and using LSD posthoc test). Each symbol [squares for mice 6-months old (6-m), diamonds for 24-months old (24-m), and circles for 28+ months old (28+-m)] indicates the test score for an individual mouse. The rectangles with error symbols (±SEM) in each grouping of ages indicates the mean value for that group.

Contractile Physiology (Figures e3-e4, Table e2)

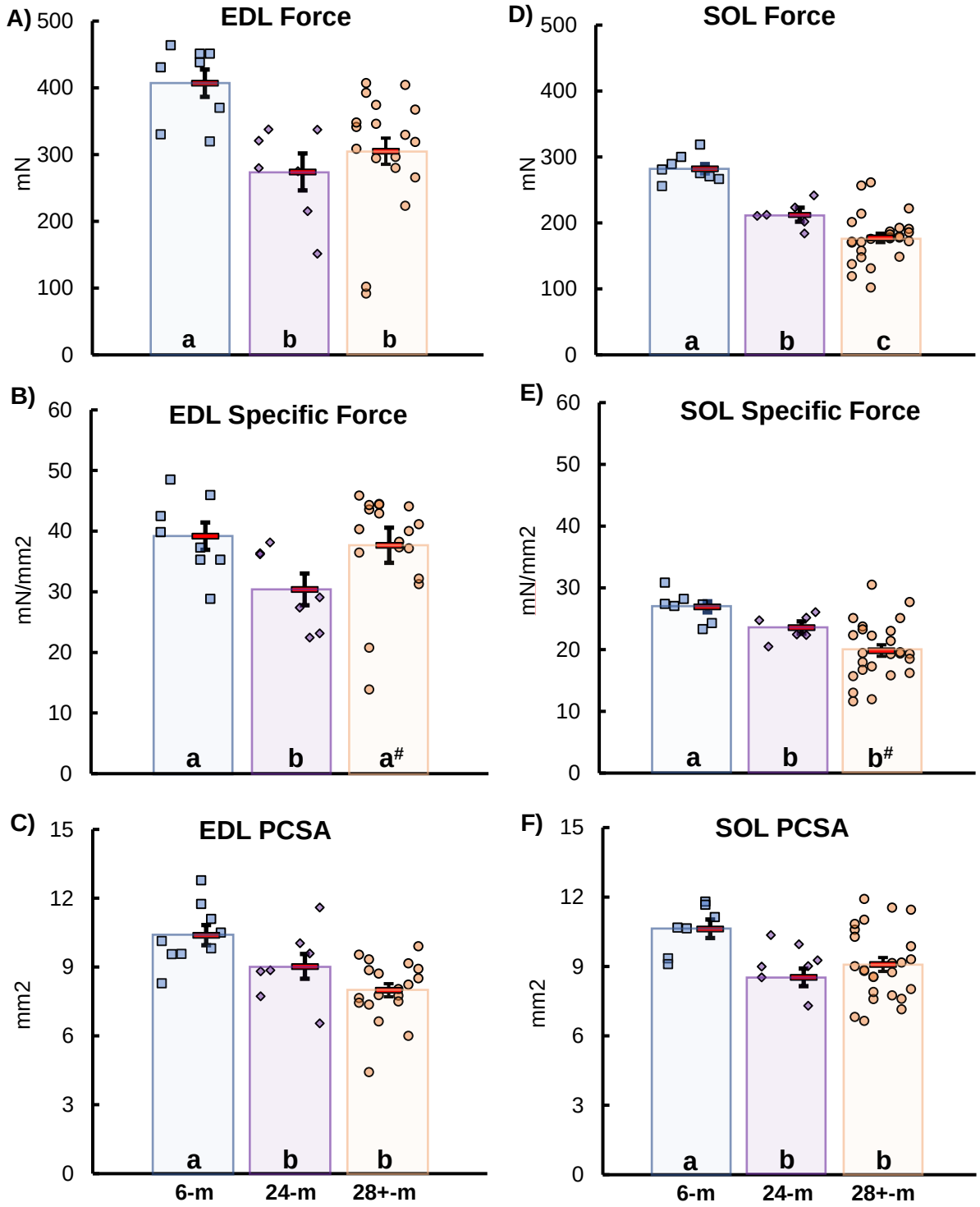

Figure e3 *ex vivo* Contractile Physiology. A) EDL P_0, B) EDL Specific Force 28-month group mean tended to be higher than 24-month**, C) EDL PCSA, D) SOL P_0, E) SOL Specific Force** 28-month group mean tended to be lower than 24-month. **F) SOL PCSA. KEY:** *Abbreviations:* EDL (extensor digitorum longus), Maximum Isometric Force (P_0), Specific Force (P_0/PCSA), PCSA (Physiological Cross-Sectional Area), SOL (Soleus), m=months of age. *Symbols:* Different letters indicate significant difference ($p<0.05$) between groups (from ANOVA using LSD posthoc test). Each symbol (squares for 6-month, diamonds for 24-month, and circles for 28+-month old mice) indicates the value for an individual mouse. The rectangles with error symbols (±SEM) in each grouping of ages indicates the mean. # indicates a trend: $0.05<p<0.10$.

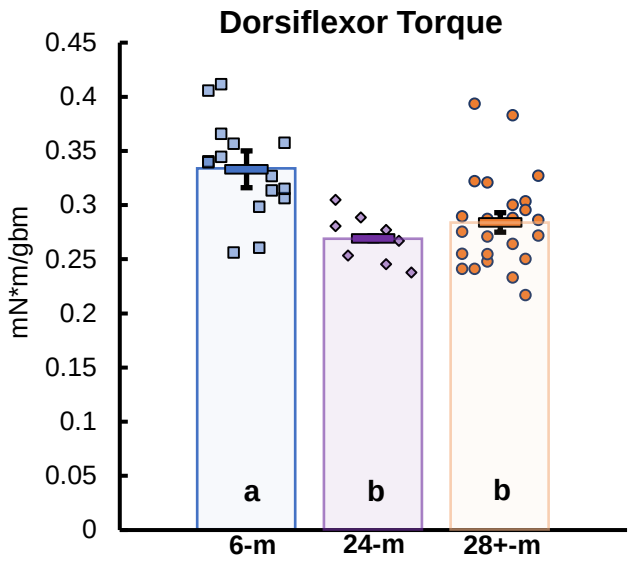

Figure e4 *in vivo* Contractile Physiology: Dorsiflexor Torque KEY: *Abbreviations:* mN=milliNewtons, m=meters, gbm=grams body mass. *Symbols:* Different letters indicate significant difference ($p<0.05$) between groups (from ANOVA using LSD posthoc test). Each symbol (squares for 6-month, diamonds for 24-month, and circles for 28+-month old mice) indicates the value for an individual mouse. The rectangles with error symbols (±SEM) in each grouping of ages indicates the mean. # indicates a trend: $0.05<p<0.10$.

		6-Months		24-months		28-months	
	n	11		8		26	
	unit	mean	SD	mean	SD	mean	SD
mean age	m	7.0	0.0	25.0	0.0	28.9	0.6
body mass	g	32.1	2.1	33.6	1.6	30.8	3.1
EDL mass	mg	13.6	2.0	**11.1**	1.7	**10.5**	2.0
EDL PCSA	mm_2	10.4	1.3	9.0	1.6	**8.0**	1.3
EDL P_t	mN	85.0	13.2	**65.5**	18.0	**67.5**	19.3
EDL Length	mm	12.3	1.6	11.7	0.9	12.2	1.1
EDL P_0	mN	407.0	57.9	**274.1**	69.2	**305.1**	88.1
EDL SpF	mN/mm^2	39.2	6.4	**30.4**	6.5	37.7	8.6
SOL mass	mg	12.9	1.3	**10.6**	1.2	**9.7**	1.7
SOL PCSA	mm_2	10.6	1.1	**9.0**	1.0	**9.1**	1.5
SOL Pt	mN	46.9	4.3	40.3	13.9	**36.0**	8.3
SOL Length	mm	11.2	0.9	11.2	1.0	**10.1**	1.1
SOL P_0	mN	282.2	20.1	**212.5**	19.6	***177.3***	36.2
SOL SpF	mN/mm_2	26.9	2.1	**23.6**	2.1	**19.8**	4.7
GAS	mg	177.2	16.1	**160.8**	10.3	***120.1***	15.9
Plant	mg	22.3	2.6	20.5	2.0	***17.4***	3.7
TA	mg	55.9	4.4	**50.3**	4.4	***42.0***	3.7
Heart	mg	171.9	17.3	**209.0**	18.3	**193.7**	33.5

Table e2 *in vitro* Muscle Physiology Subset Data **Abbreviations: PCSA** (physiological cross-sectional area), **P_t** (peak isometric twitch force), **P_0** (peak isometric tetanic force), **EDL** (extensor digitorum longus), **SOL** (soleus), **GAS** (gastrocnemius), **Plant** (plantaris), TA (tibialis anterior), **SD** (standard deviation), values in **bold** indicate significant difference ($p<0.05$) from adult values, *italics* in elderly values indicates significant difference from older values.

Regression Analysis

Correlations and Regressions: (Figures e5 – e18)

We examined the relations between the variables with regression analysis (see below for details). Using the combined set of all mice of the three ages, we examined the relationship between the determinants of CFAB using linear regression and determined there was no collinearity in the five determinants of the CFAB score. Furthermore, we determined that CFAB is predictive of age group based on logistic regression.

The variance inflation factor (VIF) for the CFAB multilinear regression of the 5 determinants (standardized) showed no evidence of multi-collinearity in the model [CFAB = (constant 95% confidence interval -0.001 to 0.000) + 0.262*grip(N) + 0.456*cling time(log10sec) + 0.213*VWR(km/day) + 0.262*treadmill(sec) + 0.306*rotarod(sec), R=1.0, p<0.001; VIF was: VWR, 1.742; rotarod, 1.519; grip strength, 1.530; treadmill, 1.095, and inverted cling, 1.245]. In the linear regressions of the individual determinants (not standardized) statistically significant moderate associations of R>0.500 or better were observed only in VWR vs. rotarod (R=0.540, p<0.001).

Using logistical regression [model: -1.892 * CFAB – 6.822 = Age (6 or 28 months)], there was statistical evidence that we could determine the age group of a mouse given its CFAB score with an overall 92.2% level of accuracy (6 month 92.3%, 28+ month 92.0%) when comparing between 6 month and 28+ month old mice (Chi-square 56.971, p<0.001). Similar results were obtained when comparing 6-month old mice to 24-month old mice [78.4% overall correct classification, 80.8% for the 6-month and 76% for 24-month; Chi-square 27.219, p<0.001; model: Age = -1.608 - 0.721 * CFAB]. The detection ability of CFAB to discern age groups also extended to determining which mouse was 24 months and which was 28 months old, though was markedly less accurate [66.0% overall correct classification, 68.0% for 24 month and 64.0% for 28 month; Chi-square 13.653, p<0.001; model -2.509 – 0.440 * CFAB = age (24 or 28 months)]. Thus, there is clear distinction and further evidence of a decay of function with age that can be chronicled using the CFAB score.

CFAB and normalized dorsiflexor torque shared a somewhat moderate linear relationship [CFAB = 40.75 * torque (mN*m / gram body mass) – 16.82, R=0.545, p=0.002, Figure e8B]. Total muscle mass (sum of SOL, EDL, TA, Plant, and GAS) was significantly different between the groups (adult: n=10, mean = 281.0 ± 7.35; older: n=10, mean = 255.7 ± 5.1, 9.0% lower than adult p=0.010; elderly: n=26, mean = 179.5 ± 3.7, 36.1% lower than adult p<0.001, 29.8% lower than older p<0.001), The linear relationship between total muscle mass and CFAB [CFAB = 0.05 * Total Muscle Mass (mg) – 15.63 , R=0.671 , p<0.001, Figure e8A] indicates that muscle mass alone accounts for 45% (R^2 = 0.450) of the variance. Normalized dorsiflexor Torque and total muscle mass were moderately correlated (linear regression: Figure e9, Normalized Dorsiflexor Torque = 0.001 * Total Muscle Mass + 0.169, R=0.600, R^2=0.340, p<0.001). If torque output was added to muscle mass in a multiple linear equation to determine CFAB there was no significantly increase of predictive value.

Power Analyses and Effect Sizes:

The effect size [$(mean_{6\text{-month}} – mean_{28+\text{-month}})$/standard $deviation_{pooled}$; standard $deviation_{pooled} = ((SD_{6\text{-month}}^2 + SD_{28+\text{-month}}^2) / 2)^{0.5}$ achieved from our experimental data comparing means of the 6-month and 28+-month groups for VWR, rotarod, grip test, treadmill, and inverted cling was 1.52, 2.1, 1.64, 0.67, and 0.96, respectively; and CFAB had an effect size of 2.21 [for comparison: an effect size of ~1.0 is considered a rather strong effect, where 84% of the experimental group is below the average of the control group, and representing only a 55% overlap between the groups, an effect size of 2 indicates 98% of experimental group would be below the control (6-7). Note that generally an effect size of 0.4 is considered "strong"; interpreted that the average experimental group member is 0.4 SD away from the mean of the control group. This higher effect size demonstrates that CFAB was able to discern differences between groups better than the individual determinants alone.

Then, using G+Power (version 3.1.9.7), we conducted power analyses using the data we collected for the 6-month and 28+-month old group means and standard deviations (sd) to calculate effect sizes and then computed how many mice were needed *per group* to produce 80% minimum power with the alpha error probability equal of 0.05 using an independent samples t-test: VWR (n=8, ES 1.52, actual power = 0.81), rotarod (n=5, ES 2.05, actual power = 0.81), grip test (n=8, ES 1.55, actual power = 0.82), treadmill (n=36, ES 0.68, actual power = 0.81), inverted cling (n=18, ES 0.97, actual power = 0.80), and CFAB (n=5, ES 2.28, actual power = 0.88). Taken together these results demonstrate that CFAB overall produces a higher effect size with smaller numbers of animals at an increased power

as the individual determinants alone, indicating that CFAB is a successful composite scoring system that improves detectability of differences between groups.

Linear Regressions of the CFAB Determinants (Figures e5-e7): Note for all regression graphs: The dashed line is representative of the simple linear regression equation given in each graph for the combined age groups..

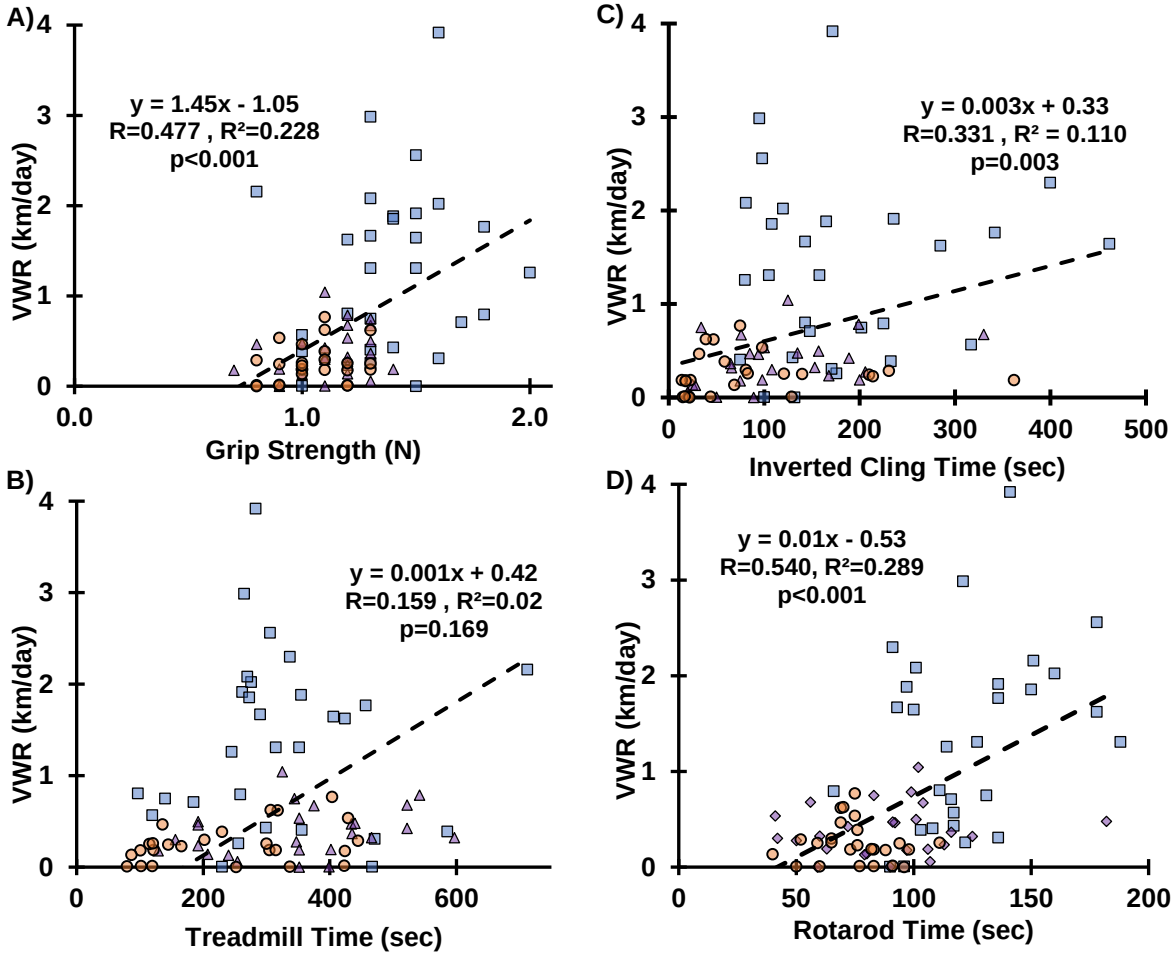

Figure e5 Linear Regressions: A) VWR vs. Grip Strength, B) Treadmill vs. VWR. C) VWR vs. Inverted Cling, D) VWR vs. Rotarod. *Abbreviations:* VWR=voluntary wheel running, km=kilometers, N=newtons, sec=seconds *Symbols:* Each symbol equals the regression of the scores of an individual mouse. Squares=6-month old mice, Triangles=24-month old mice, and circles=28+-month old mice. Dashed line and equation equal the simple linear regression of the combination of all three groups.

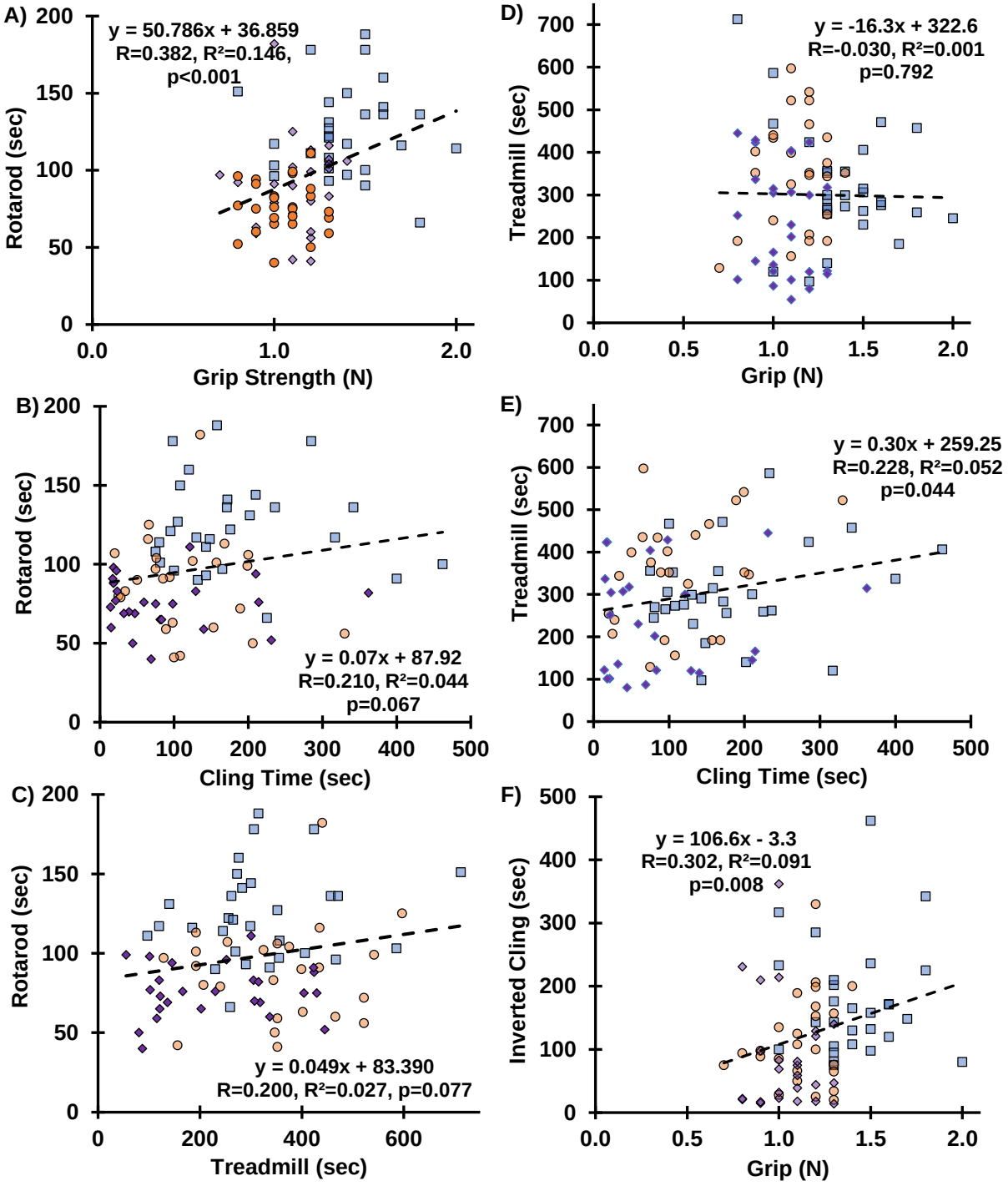

Figure e6 Linear Regressions: A) Rotarod vs. Grip B) Rotarod vs. Inverted Cling, C) Rotarod vs. Treadmill, D)Treadmill vs. Grip, E) Treadmill vs. Inverted Cling. F) Inverted Cling vs. Grip *Abbreviations:* sec=seconds, N=Newtons *Symbols:* Each symbol equals the regression of the scores of an individual mouse. Squares=6-month old mice, Triangles=24-month old mice, and circles=28+-month old mice. Dashed line and equation equal the simple linear regression of the combination of all three groups.

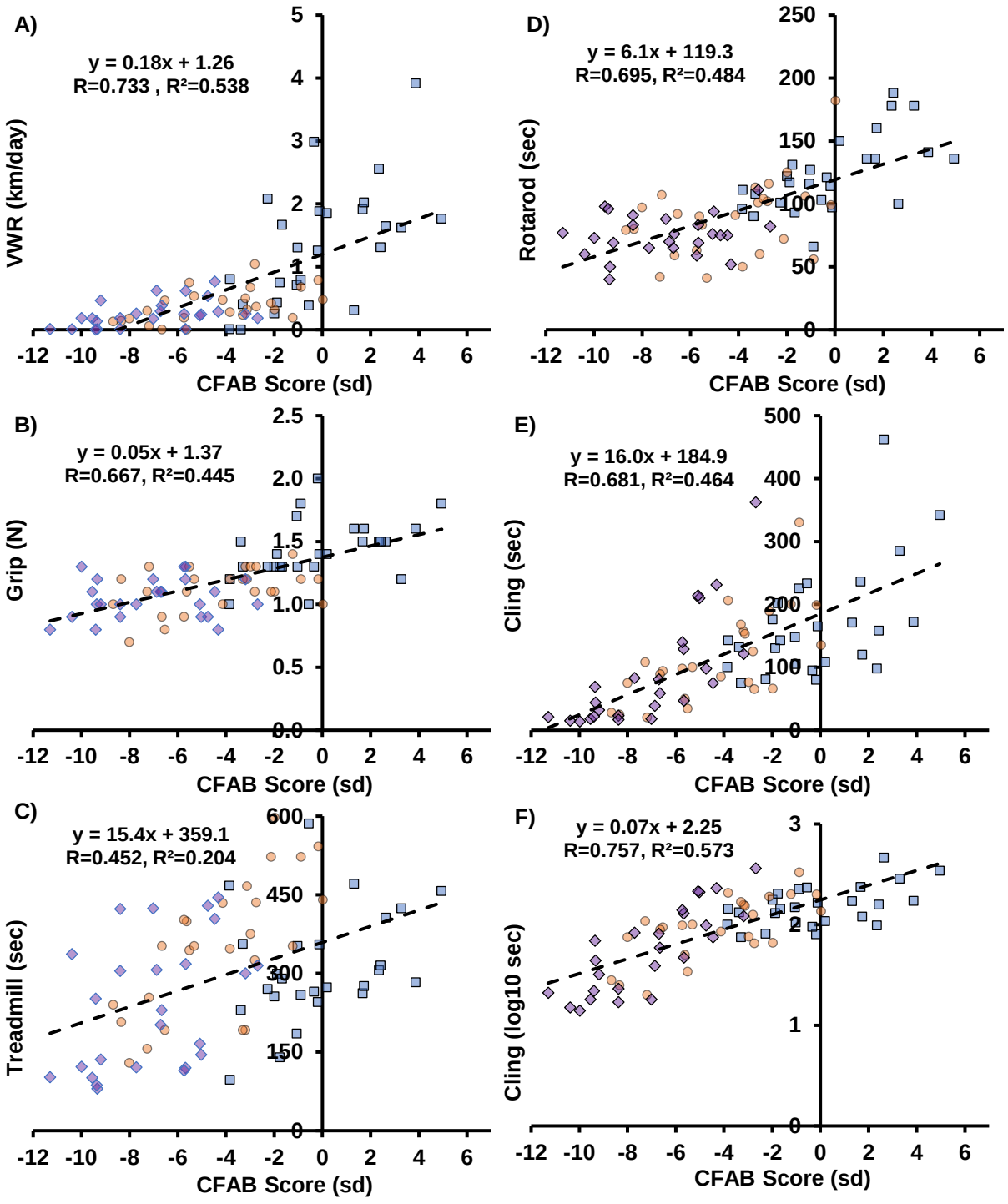

Figure e7 Linear Regressions: A) VWR vs. CFAB, B) Grip vs. CFAB, C) Treadmill vs. CFAB, D) Rotarod vs. CFAB, E) Inverted Cling vs. CFAB, F) Inverted Cling (log10) vs. CFAB *Abbreviations:* km=kilometers, N=newtons *Symbols:* Each symbol equals the regression of the scores of an individual mouse. Squares=6-month old mice, Triangles=24-month old mice, and circles=28+-month old mice. Dashed line and equation equal the simple linear regression of the combination of all three groups.

Dorsiflexor Strength Related to Total Muscle Mass (Figure e8)

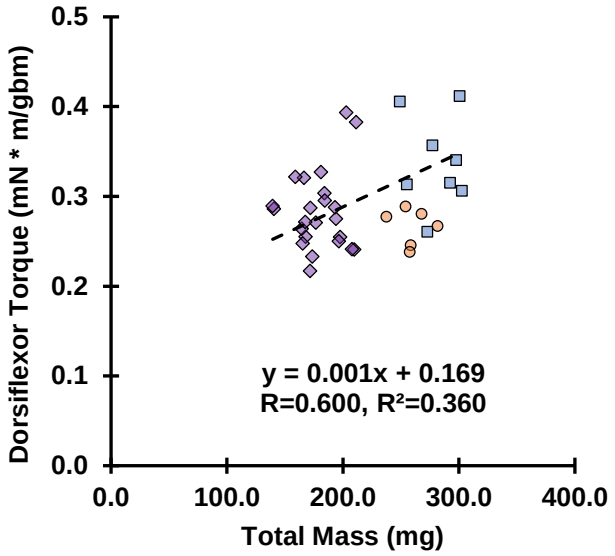

Figure e8 Dorsiflexor Torque vs. Total Muscle Mass. Total mass (mg) is the sum of the mass of the plantaris, soleus, EDL, tibialis anterior, and gastrocnemius muscles. *Abbreviations:* m=meters, mN=milliNewtons, gbm=gram body mass, mg=milligrams *Symbols:* Each symbol equals the regression of the scores of an individual mouse. Squares=6-month old mice, Triangles=24-month old mice, and circles=28+-month old mice. Dashed line and equation equal the simple linear regression of the combination of all three groups.

Table e3 CFAB Determinant Correlation Matrix: numbers=Pearson's R, from simple linear regression, m=months of age

		Rota				Treadmill				Grip Test				Cling			
		6-m	24-m	28+-m	all	6-m	24-m	28+-m	all	6-m	24-m	28+-m	all	6-m	24-m	28+-m	all
VWR	6-m	0.349				0.058				0.107				0.070			
	24-m		0.051				0.288				0.220				0.339		
	28+-m			-0.094				0.342				0.238				0.046	
	all				0.539				0.159				0.477				0.331
Rota	6-m	1.000				0.131				0.017				0.089			
	24-m		1.000				0.108				0.014				0.157		
	28+-m			1.000				0.094				0.039				<0.001	
	all				1.000				0.200				0.381				0.210
Tread	6-m	0.131				1.000				0.322				0.211			
	24-m		0.108				1.000				0.200				0.346		
	28+-m			0.094				1.000				0.233				0.235	
	all				0.200				1.000				-0.030				0.228
Grip	6-m	0.017				0.322				1.000				0.007			
	24-m		0.014				0.200				1.000				0.197		
	28+-m			0.039				0.233				1.000				0.134	
	all				0.381				-0.030				1.000				0.302
Cling	6-m	0.089				0.211				0.007				1.000			
	24-m		0.157				0.346				0.197				1.000		
	28+-m			<0.001				0.235				0.134				1.000	
	all				0.210				0.302				0.540				1.000
CFAB	6-m	0.626				0.335				0.445				0.540			
	24-m		0.415				0.698				0.405				0.679		
	28+-m			0.205				0.424				0.164				0.759	
	all				0.695				0.452				0.667				0.681

		VWR			
		6-m	24-m	28+-m	all
CFAB	6-m	0.615			
	24-m		0.528		
	28+-m			0.464	
	all				0.733

eDISCUSSION:

Comparison of CFAB to Frailty Phenotype and NMHSSC

Frailty is the consequence of the failure of the body to maintain homeostasis and can manifest in many of the same detriments associated with sarcopenia: poor prognosis after procedures, loss of functional ability leading to inadequate performance of activities of daily living, eventual loss of independence, and an increased mortality rate. Clinically there are two well established methods for diagnosing frailty: the Fried Frailty Phenotype and the Rockwood Deficit Accumulation Frailty Index (8-9). There is a body of previous work that has reverse-translated both the Fried Phenotype and the Rockwood Deficit Accumulation Index into the C57BL/6 mouse model (1,10-16). Using similar functional tests as CFAB, the most recent iteration of the mouse frailty phenotype measured the 5 components inherent in the human clinical frailty phenotype score using well-validated mouse physical function tests: body weight (mass at 23 months), endurance (treadmill), walking speed (rotarod), strength (grip test), and physical activity (running wheel). In this model, mice are ranked by performance within the cohort at the age of 23 months where the bottom quintile in each test is considered below the cut-off (except for mass which considers the top quintile to be associated with frailty) (11,12). If three or more of the determinants are below the cutoff threshold, the mouse is considered frail. Instead of determining cut-off points within a cohort, CFAB produces a numerical summarization of its 5 determinants based upon a young cohort to determine the effect of age on athletic and functional performance.

To determine whether poor performing mice on the CFAB scale also correlate with frailty, we used the data collected in our three ages and then applied the Frailty Phenotype system (18,22) to determine which animals would be deemed frail: cutoffs were set to be at the bottom 20% of the 24-month old group (n=24) of VWR (0.147 km/day), grip strength (1.0 N), treadmill (195.4 sec), rotarod (59.2 sec), inverted cling (1.72 log10_ sec), and taking body mass into account (upper 20%, 35.5g). Any mouse that fell below the cutoff in 3 or more of the categories is considered frail (2 considered pre-frail).

Bauman and colleagues (11) used grip strength rather than the inverted cling used in the original Frailty Phenotype (1,16). We assessed our mice using either or both (counted only once towards the frailty phenotype if a mouse was below the threshold in both). In the older group, there was no overlap: the mice that scored under the cutoff in the grip test and the inverted cling were not the same individuals. This makes sense because we know that in our study the correlation between grip strength and inverted cling was slight ($R=0.302$, $p=0.008$), thus emphasizing that grip strength and inverted cling a measuring two different parameters of physical strength. Notably in the elderly group, the inverted cling detected more mice (n=12) than the grip strength meter (n=7) to be under the threshold, with only 4 overlaps (mice below the threshold in both tests). One potential mechanism for this disparity in detection may be that inverted cling has a more nuanced sensitivity with a wider possible range because the grip meter is limited to 0.1 N detectability.

In the older group of mice using the grip strength test, 3 mice were deemed frail, a result (13 %) close to the ratio found in Bauman and colleagues (11) (~11%). If the cling was used, 2 mice were deemed frail, with 4 deemed frail and 3 pre-frail if we counted both the cling and grip meter. In the 28+ group, 21% in our model (if we count either grip or cling) were frail (10 were pre-frail) compared to 33.3% frailty in the 29-month old group in the previous work (11). When using the grip test alone only 17% are frail. However, the bottom 20% of performers in CFAB (< -6.277) were not all frail, as CFAB measures the effect age on function, not frailty per se. All frail mice as measured by the Frailty Phenotype were deemed in the lower quintile of CFAB. In the older subset of mice all four frail mice were in the CFAB bottom quintile, but one performed 5th worse in CFAB while not deemed frail or prefrail using the Frailty Phenotype score. In the elderly subset, all frail mice and 50% of the pre-frail mice were under the CFAB threshold (n=10 mice under the threshold altogether). Overall, the evidence supports that CFAB may be used as a proxy for detecting frailty and for noting some pre-frail individuals.

The neuromuscular healthspan scoring system (2) used *in vitro* contractile isometric force (P_0) output of the EDL, rotarod and the inverted cling test to measure neuromuscular decline while accounting for individual variability with multi-linear regression. In our subset of mice that underwent EDL contractile testing and for which we had all needed data, we used the equations from the earlier paper (2) (but with data from the entire cohort of this study) and determined that our mice had NMHSS of 2.8, 3.8, and 3.4, respectively for adults (n=8), older adults (n=5) and elderly (n=16). One caveat is that the NMHSS used the average of all trials and the current study reports the maximum value, so this comparison is faulty. Note: a mouse that is exactly average for its age group and whose predicted score equals the actual score on all tests would be scored a 3. Thus, the 5 mice tested in the older adult group seemed to be

extraordinarily healthy compared to their peers as measured in the original study, with much of this variation being driven by exceptional scores on the rotarod in this small cohort (2).

	NMHSS (1)	FI (2)	FIAV (3)	CFAB (4)
Number of Unique Determinants	3	3	3	5
Includes Invasive Techniques	Yes	No	No	No
Continuous Data?	Yes	No	Yes	Yes
Reliance on Multiple Linear Regression?	Yes	No	No	No
Improves Power vs. Individual Determinants	Yes	N/A	Yes	Yes
Uses Reference Group	No	No	No	Yes

Table e4 Comparison of CFAB, FI, FIAV, NMHSS. Abbreviations: NMHSS, NeuroMuscular Healthspan Scoring System; FI, Frailty Phenotype Index; FIAV, Frailty Intervention Assessment Value; CFAB, Comprehensive Functional Assessment Battery. 1) from doi: 10.1093/gerona/glt032; 2) from doi: 10.1093/gerona/glt188; 3) from doi: 10.1093/gerona/glu163; 4) from current study.

Frailty Invention Assessment Value (FIAV) and CFAB:

In previous work we detailed the predecessor to CFAB, the Frailty Invention Assessment Value (FIAV) (1). This composite scoring system used rotarod maximum speed at fall (revolutions per minute), inverted cling grip test (seconds), activity rate (voluntary wheel running in km/day), and a derived endurance score calculated from two of the other components (time on rotarod + time on inverted cling)/2) as its four determinants. Therefore, unlike CFAB which has a dedicated endurance measurement (treadmill time), FIAV makes a combined derived score reusing two of its other determinants (rotarod and inverted cling) giving those two tests a greater influence on the score. This was admittedly a weakness of the previous system, which has been improved by the introduction of the treadmill as an independent measurement in CFAB. The FIAV is very similar to CFAB in that it counts the number of standard deviations an individual mouse's score deviates from the baseline mean but uses fewer unique tests. However, in FIAV, one test (FIAV1) is given pre-intervention and the second test (FIAV2) is given post-intervention, by definition FIAV = FIAV2 – FIAV1, with the standardized score dependent upon the mean of the cohort itself (not a reference group as in CFAB). By inference this implies that since CFAB uses a similar system to score its determinants as FIAV (with the addition of two new validated tests, which are shown to be independent and thus measuring different aspects of function) it would be able to discern changes just like FIAV if used in such a manner (CFAB1-CFAB2=change due to longitudinal condition or treatment. Thus, if a cohort were to be measured using the CFAB determinants both pre-intervention and then again post-intervention, the difference in scores would be the "CFAB FIAV". However, instead of subtracting the sum of standardized scores pre- and post-treatment to get the difference as in FIAV, we might simply compare the pre-intervention CFAB to the post-intervention CFAB with an ANOVA or t-test. We calculated the FIAV1 of the three age groups in the current study and compare them to CFAB in Table e5. Please note that CFAB is standardized based upon the control group (the 6-month old group), whereas FIAV1 is calculated based upon the mean and standard deviation within the cohort, and as such they are truly not comparable in any way other than in the 6-month old group in which the determinants are calculated similarly (hence the low R^2 from linear regression in the older groups and a moderate correlation in the 6-month old group).

Age Group	CFAB	FIAV1	r^2	R
6-month	-0.13	1.51	0.65	0.81
24-month	-4.24	-2.22	0.06	0.24
28+-month	-7.21	-3.46	0.05	0.22

Table e5 FIAV1 and CFAB. Using data from this study we recreated the FIAV1 as described in Graber, et al. 2014 and ran a linear regression with CFAB (R and r^2 are shown). The 6-month group is the only valid comparison due to methodology differences (see text) in the two systems. The six-month group has a strong correlation between the two measurements indicating that they are measuring function similarly.

References

1. Graber TG*, Ferguson-Stegall L*, Liu H, Thompson LV. Voluntary Aerobic Exercise Reverses Frailty in Old Mice. The Journals of Gerontology. Series A, Biological Sciences. 2015;70(9): 1045-1058. doi: 10.1093/gerona/glu163 Epub ahead of print: 09/30/14.
2. Graber TG, Ferguson-Stegall L, Kim J-H, Thompson LV. C57BL/6 Neuromuscular Healthspan Scoring System. The Journals of Gerontology. Series A, Biological Sciences. 2013;68(11): 1326-1336. doi: 10.1093/gerona/glt032
3. Graber TG, Kim J-H, McLoon LK, Grange RW, Thompson LV. C57BL/6 Lifespan Study: Age-Related Declines in Muscle Power Production and Contractile Velocity. Age. 2015;37: 36-. doi: 10.1007/s11357-015-9773-1
4. Neelakantan H*, Brightwell CR*, Graber TG, Maroto R, Wang H-Y L, McHardy SF, Papaconstantinou J, Fry CS, Watowich SJ. Small molecule nicotinamide N-methyltransferase inhibitors activate senescent muscle stem cells and improve regenerative capacity of aged skeletal muscle. Biochem Pharmacol. 2019;May;163:481-492. doi: 10.1016/j.bcp.2019.02.008. Epub 2019 Feb 10
5. Graber TG, Rawls BL, Tian B, Durham WJ, Brightwell CR, Brasier AR, Rasmussen BB, Fry CS. Repetitive TLR-3-mediated Lung Damage Induces Skeletal Muscle Adaptations and Cachexia. Exp Gerontol. 2018; Feb 13. pii: S0531-5565(17)30667-8. doi: 10.1016/j.exger.2018.02.002. [Epub ahead of print]
6. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. J Grad Med Educ. 2012;4(3):279–282. doi:10.4300/JGME-D-12-00156.1
7. Coe R. It's the effect size, stupid: what "effect size" is and why it is important. Paper presented at the 2002 Annual Conference of the British Educational Research Association, University of Exeter, Exeter, Devon, England, September 12–14, 2002. Accessed via WWW on 10/04/19 at: http://www.leeds.ac.uk/educol/documents/00002182.htm.
8. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA, and Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001; 56:M146–56. doi: 10.1093/gerona/56.3.M146
9. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. Sci World J. 2001; 1:323–36. doi: 10.1100/tsw.2001.58
10. Baumann CW, Kwak D, Thompson LV. Assessing onset, prevalence and survival in mice using a frailty phenotype. Aging (Albany NY). 2018;10(12):4042–4053. doi:10.18632/aging.101692
11. Baumann CW, Kwak D, Thompson LV. Assessing onset, prevalence and survival in mice using a frailty phenotype. Aging (Albany NY). 2018;10(12):4042–4053. doi:10.18632/aging.101692
12. Kwak D, Baumann CW, Thompson LV. Identifying Characteristics of Frailty in Female Mice using a Phenotype Assessment Tool. J Gerontol A Biol Sci Med Sci. 2019 Apr 8. pii: glz092. doi: 10.1093/gerona/glz092. [Epub ahead of print]
13. Parks RJ, Fares E, Macdonald JK, Ernst MC, Sinal CJ, Rockwood K, Howlett SE. A procedure for creating a frailty index based on deficit accumulation in aging mice. J Gerontol A Biol Sci Med Sci. 2012; 67:217–27. doi: 10.1093/gerona/glr193
14. Whitehead JC, Hildebrand BA, Sun M, Rockwood MR, Rose RA, Rockwood K, Howlett SE. A clinical frailty index in aging mice: comparisons with frailty index data in humans. J Gerontol A Biol Sci Med Sci. 2014; 69:621–32. doi: 10.1093/gerona/glt1
15. Rockwood K, Blodgett JM, Theou O, Sun MH, Feridooni HA, Mitnitski A, Rose RA, Godin J, Gregson E, Howlett SE. A frailty index based on deficit accumulation quantifies mortality risk in humans and in mice. Sci Rep. 2017; 7:43068. doi: 10.1038/srep43068
16. Liu H*, Graber TG*, Ferguson-Stegall L, Thompson LV. Clinically Relevant Frailty Index for Mice. The Journals of Gerontology. Series A, Biological Sciences. 2014;69(12): 1485-1491. doi: 10.1093/gerona/glt188 Epub ahead of print: 12/09/13.